27.8 A 128×128 33mW 30frames/s Single-Chip Stereo Imager

Ralf M. Philipp, Ralph Etienne-Cummings

Johns Hopkins University, Baltimore, MD

A single chip containing two 128×128 pixel imagers and current-mode disparity computation circuitry is described. Previous stereo vision sensors have not integrated computation circuitry [1] on the focal plane, or been limited to single depth values [2]. The previous single-chip stereo imager [3] used continuous-time imaging and computation circuits. These circuits were non-linear, power hungry, and had high levels of mismatch. They limited the operation of the previous chip, which was never equipped with optics to be able see an actual stereo pair, to very high-contrast scenes. The improvements of the chip described in this work include a more accurate computation block, the construction of optics, a 50% reduction in power consumption, and a linear current-mode imager with an uncorrected FPN that is lower than what has been reported for similar imagers [4].

The algorithm described in [3], a simplified version of block matching using the sum of absolute difference (SAD) matching metric, is implemented on the chip. The output of the chip is the disparity Δv between the left and right images of the stereo pair. For coplanar imagers (looking in the same direction) Δv maps to depth z by $z = bf / p\Delta v$ (where b: baseline, f: lens focal length, p: pixel pitch). Figure 27.8.1 shows a system diagram for the chip. For each location in the right image, the SAD metric for every candidate disparity in the left image is computed in parallel. The serial-scan parallel-computation architecture allows the chip to produce full-frame disparity maps (114×125 pixels) at 30frames/s while clocking the computation circuits at 430kHz. The chip also provides left and right image outputs and a spatial highpass of the right image, a match-validity-metric (MVM). The chip has a digital interface (excluding the image outputs) and can be used as a distance sensor without any external ADCs, DACs, or amplifiers.

The integrating current-mode active pixel sensor (APS) implemented in the chip (Fig. 27.8.1) provides a linear conversion from light to output current [4]. The g_m from the integrating node V_{pix} to output current I_{pix} is linearly adjustable by varying V_{col} (3V nominal) as long as M2 is kept in the triode region. V_{reset} is set at approximately $|V_{tP}| \approx 0.8V$ below V_{DD}, ensuring that M2 is always above threshold. Selecting multiple rows provides a vertical averaging (summation) of pixel values, as required by the algorithm. The imagers exhibit an uncorrected FPN of 1.2% (versus 1.9% and 0.8% before and after CDS in [4]). Readout circuits produce over half of this mismatch; the pixel FPN is calculated to be 0.9% which is low for a small (10μm pitch) and uncorrected current-mode pixel.

V_{col} is fixed using column-parallel first-generation current-conveyors (CCI). Two copies are made of the column currents I_{col}: one for disparity computation and one for the image output. The image and disparity outputs can be scanned column-wise independently. Computing the SAD metric over the disparity search-space requires 3 operations: addition, subtraction, and absolute value (full wave rectification). Analog current-mode processing provides addition without any extra circuitry, however, subtraction requires a current mirror to implement the sign change. The chip contains 1713 instances of the synchronous full-wave current rectifier as shown in Fig. 27.8.2 [5]; it occupies 612μm² and a has a 2μA comparator bias current. It requires <300ns for a complete operational cycle on input currents of ±0.1 to 10μA.

The final mathematical operation required to implement the algorithm is "arg min"; a synchronous loser-take-all (LTA) [6] finds the location of the smallest of the 114 computed SAD currents. The LTA, shown in Fig. 27.8.3, is a 128-input 7-level binary tree of 2-input unit cells. The LTA also provides 1 of the 2 on-chip MVMs: the externally controllable "*max SAD*" input shown in Fig. 27.8.1 (replicated 14 times to balance the LTA tree). Low quality matches (i.e. with high SAD values) are rejected in favor of the *max SAD* input. 128 current inputs of the LTA also serve as an active-low one-hot digital output that is encoded and provided as a 7b parallel digital disparity output.

The optics are shown in Fig. 27.8.4. A right-angle prism mirror separates the left and right images. The minimum lens to prism separation is determined by the depth-of-field of the lens, the angular separation (3.3° at f=4mm) of the imagers, and the corner bevel of the prism mirror. Mirrors on adjustable mounts are then used to align the left and right images to produce a coplanar image pair. After initial positioning, the outputs of the chip are used for mirror alignment. Vertical misalignment is detected by the absence of consistent disparity outputs. Horizontal alignment is achieved by aiming the mirrors until targets at calibrated depths have correct disparity values. A calibrated depth corresponding to zero disparity is infinity ($z_\infty > 2bf / p$).

Figure 27.8.5 shows the depth map of a small box and a miniature traffic cone taken from a 30frames/s sequence. The inset image is the right image after 5-pixel on-chip vertical averaging (part of the algorithm). The depth results from the chip are filtered (pass/fail) using the on-chip *max SAD* and spatial highpass MVMs, and a median filter to remove spurious points. The front face of the box is resolved at a much higher depth resolution than the cone. The difference between the nearest 2 depth values on the box (closest z=77cm) is 1.67cm, whereas the cone has 2 discrete depth values differing by 10.5cm. This decrease in resolution is quadratic with depth and is intrinsic to stereo vision (see Fig. 27.8.6). No depth values are found for the right side of the box, since it is not seen by the left imager, thus failing the *max SAD* MVM. No depth values are returned for dark areas, as they do not contain sufficient contrast to pass the spatial highpass MVM.

A summary of the chip performance is provided in Fig. 27.8.6. The chip, shown in Fig. 27.8.7, provides full frame depth maps at 30frames/s while consuming 33.2mW. Operation at 40frames/s increases power consumption to 33.6mW, but at the cost of sparser depth maps.

Acknowledgements:
The authors thank Vance Tyree at MOSIS and Yu Mike Chi for their assistance.

References:
[1] T. Kato, et al., "A Binocular CMOS Range Image Sensor with Bit-Serial Block-Parallel Interface Using Cyclic Pipelined ADC's," *Dig. Symp. VLSI Circuits*, pp. 270-271, June, 2002.
[2] M. Tanaka, et al., "Autofocus Modules with MOS Analog Sensors," *Fuji Electric Review*, vol. 45, no. 2, pp. 56-58, 1999.
[3] R.M. Philipp, R. Etienne-Cummings, "Single-Chip Stereo Imager," *Analog. Integrated Circuits & Signal Proc.*, vol. 39, pp. 237-250, June, 2004.
[4] V. Gruev, R. Etienne-Cummings, T. Horiuchi, "Linear Current Mode Imager with Low Fix Pattern Noise," *ISCAS 2004*, pp. 860-863, May, 2004.
[5] R.M. Philipp, R. Etienne-Cummings, "Low Power Current Rectifiers for Large Scale Current-Mode Signal Processing," *ISCAS 2004*, pp. 229-232, May, 2004.
[6] B.M. Wilamowski, et al., "Low Power Current Mode Loser-Take-All Circuit for Image Compression," *9th NASA Sym. on VLSI Design*, pp. 7.6.1-7.6.8, 2000.

 1-4244-0079-1/06/$20.00

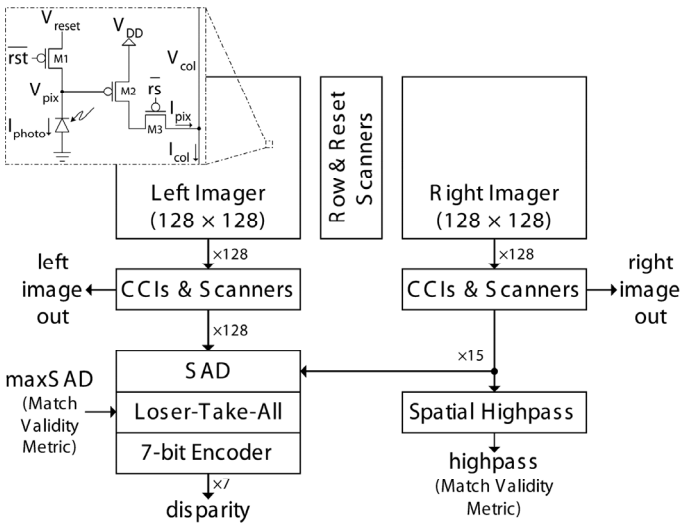

Figure 27.8.1: Chip overview with APS.

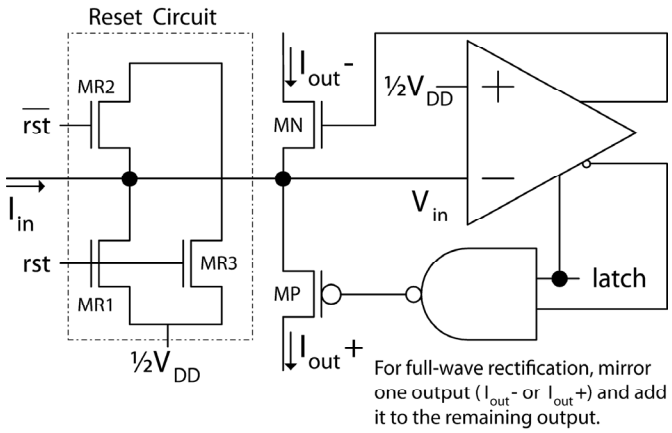

Figure 27.8.2: Full-wave current rectifier (absolute value).

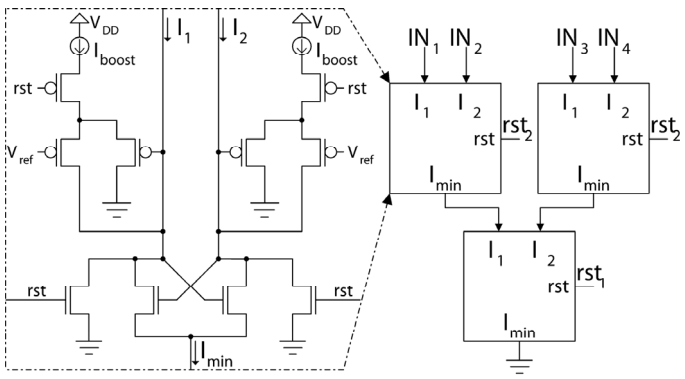

Figure 27.8.3: Loser-take-all unit cell & 2-level tree [6].

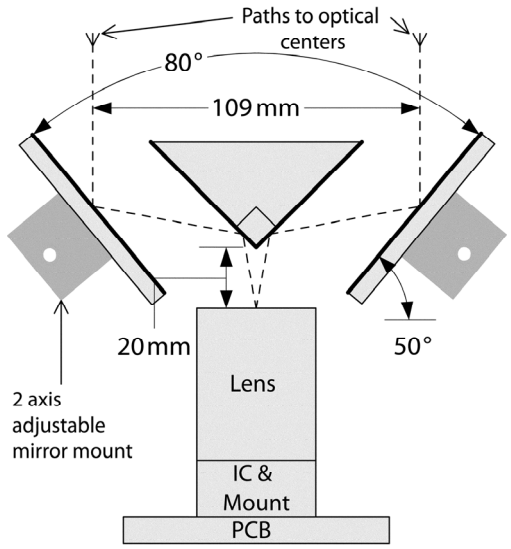

Figure 27.8.4: Optics (top view, to scale).

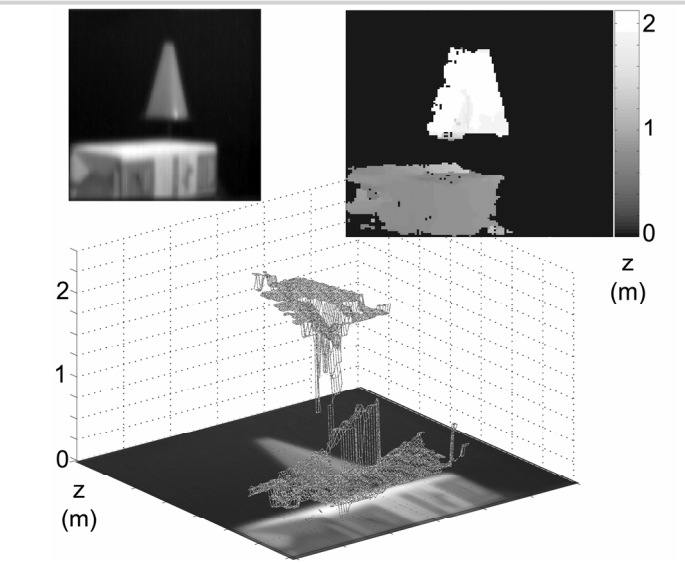

Figure 27.8.5: Depth map & right imager picture.

Technology	0.35µm CMOS (4M, 2P)
Chip area	3.5×3.3 mm^2
Operating voltage	3.3 V
Power consumption	33.6mW *(at 40 fps)*
Imager size	$2 \times 128 \times 128$ pixels
Pixel size	10×10 µm^2
Pixel fill factor	26%
FPN	1.2%
	(current-mode imaging, no offset correction)
Frame rate (nominal)	30 fps
Frame rate (maximum)	40 fps
Depth map size (x × y)	114 × 125
Minimum resolvable depth	0.39 m *(optics dependent)*
Depth	$z = bf \,/\, p\Delta v$
Depth resolution *(continuous approximation)*	$\left\lvert\frac{\partial z}{\partial \Delta v}\right\rvert \approx \frac{bf}{\Delta v^2 p} = \frac{z^2 p}{bf}$

b & *f* are optics dependent, $p = 10$µm

Figure 27.8.6: Chip performance summary.

Continued on Page 669

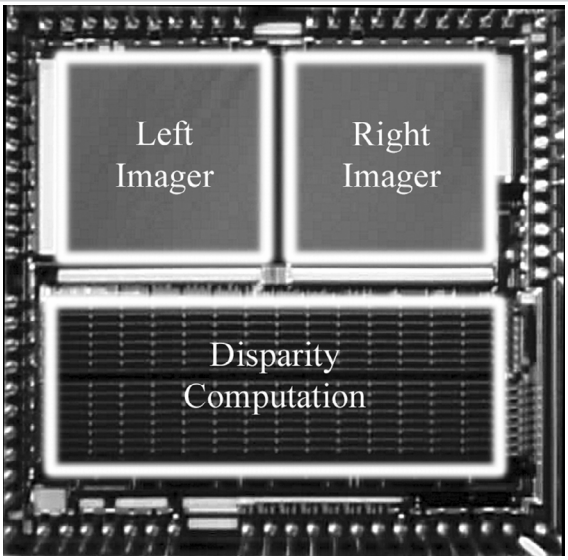

Figure 27.8.7: Micrograph.

